

Формування кластерів трудових мігрантів за ступенем ризику інфікування вірусом імунодефіциту

О. Ю. Мулеса, В. Є. Снитюк, І. С. Миронюк

Розглядається задача формування підгруп за ступенем ризику інфікування вірусом імунодефіциту у групі трудових мігрантів, як задача кластеризації. Побудовано математичну модель задачі кластеризації та виконано адаптацію методу еволюційної кластеризації об'єктів до визначення груп трудових мігрантів. Здійснено розбиття групи трудових мігрантів на кластери відповідно до ступеня ризику інфікування вірусом імунодефіциту.

Ключові слова: еволюційна кластеризація, група трудових мігрантів, ризик інфікування вірусом імунодефіциту людини

1. Вступ

В умовах концентрованої стадії епідемії ВІЛ-інфекції/СНІДу, яка наразі констатується в Україні [1], коли епідемія переважно розвивається в окремих, уразливих щодо інфікування ВІЛ групах населення [2], основні заходи протидії епідемії сфокусовані саме в середовищі представників даних груп. Зважаючи на те, що загально визнані групи підвищеного ризику (ГПР) інфікування ВІЛ є досить неоднорідними за своєю структурою [3], а також враховуючи обмеженість ресурсної бази, стверджується про необхідність розробки цілеспрямованих профілактичних інтервенцій та впливів на їх окремі, максимально однорідні за соціально-демографічними характеристиками підгрупи на засадах клієнт-центрованого підходу. Оскільки в основі такого лежить реалізація заходів, які є привабливими і зручними для представників цільової групи [4], важливою умовою його реалізації є дослідження основних соціально-демографічних характеристик таких підгруп ГПР та визначення рівня поширення практик ризикованої щодо інфікування ВІЛ поведінки в кожній підгрупі. Даний підхід дозволить визначити пріоритетність необхідності впровадження профілактичних програм для різних підгруп ГПР та сформувати максимально зручні та наближені до представників ГПР заходи програми.

Однієї із провідних ГПР інфікування ВІЛ у Закарпатській області є зовнішні та внутрішні трудові мігранти [5]. Ця група є досить багаточисельна і різноманітна за своїм складом, тому формування окремих підгруп, однорідних за соціально-демографічними ознаками та визначення ступеню ризику інфікування ВІЛ для кожної з таких підгруп в даній ГПР є необхідним і актуальним.

Задача поділу ГПР на підгрупи відповідно до соціально-демографічного портрету їх представників математично може бути представлена як задача кластеризації [6].

2. Аналіз літературних даних та постановка проблеми

Задача кластеризації полягає у розділенні основної множини на підмножини об'єктів, схожих між собою за певним критерієм [7, 8]. Існує ряд методів розв'язування задач кластеризації. Аналіз сучасних наукових публікацій, присвячених моделям і методам кластеризації об'єктів, показав, що актуальною є розробка методів кластеризації для розв'язання реальних прикладних задач. Так, в [9, 10] описані статистичні методи кластеризації, які базуються на теорії ймовірнісних розподілів та застосовуються, наприклад, адаптивних системах, що навчаються. В роботі [11] наведено нейромережний метод розв'язання задачі кластеризації та доведено його ефективність. Теоретичні основи власне нейромережних методів містяться в [12]. Такі методи вимагають від дослідників значного досвіду при налаштуванні значень основних параметрів, що ускладнює їх використання при розв'язуванні практичних задач. Методи групи деревовидної кластеризації [13] полягають у побудові дерева кластерів. Методи нечіткої кластеризації, серед яких основним методом є метод нечітких с-середніх [14], що базується на застосуванні апарату нечітких множин, дозволяють одночасно відносити один і той же об'єкт до різних кластерів з різними ступенями належності. Адаптація методу нечітких с-середніх, яка наведена в [15], дозволяє робити розбиття об'єктів на кластери з врахуванням ступенів впливу ознак на належність об'єкта кластеру.

Окрему групу утворюють евристичні методи кластеризації. Так в [16] представлено евристичний метод кластеризації, який базується на аналізі матриць відмінностей. Робота [17] присвячена розробці гібридних евристик для розв'язання задачі кластеризації. Задача кластеризації часових рядів та методи її розв'язання, які базуються на оцінці ступенів невизначеності параметрів авторегресійних моделей, приведені в [18]. Методам автоматичної кластеризації, які базуються на методі опорних векторів присвячені роботи [19, 20].

Точність розв'язання задачі кластеризації залежить від розподілу об'єктів по області дослідження, кількості необхідних ітерацій для формування та уточнення кластерів, інших проблем, пов'язаних із особливостями застосування різних методів.

Враховуючи важливість та актуальність проблеми визначення ризиків інфікування ВІЛ окремими групами населення, доцільною є розробка таких моделей і методів кластеризації, які б дозволили її ефективно вирішувати.

3. Ціль та задачі дослідження

Метою дослідження є підвищення ефективності процесів прийняття рішень при визначенні кластерів трудових мігрантів за ознакою інфікування вірусом імунодефіциту людини шляхом розробки моделей і методів еволюційної кластеризації.

В межах зазначеної мети було поставлено такі задачі:

- побудувати математичну модель задачі кластеризації та виконати адаптацію методу еволюційної кластеризації об'єктів до визначення груп трудових мігрантів;

– виконати експериментальну верифікацію розробленого методу для задачі визначення груп осіб за соціально-демографічними ознаками в ГПР – трудових мігрантів;

– здійснити порівняльний аналіз результатів розв’язання задачі кластеризації заданої ГПР різними методами кластеризації.

4. Адаптація методу еволюційної кластеризації до формування груп трудових мігрантів

Розглянемо задачу кластеризації об’єктів у такій постановці. Нехай дано множину об’єктів $O = \{O_1, O_2, \dots, O_n\}$, кожен з яких характеризується за множиною ознак $X = \{X_1, X_2, \dots, X_m\}$. Нехай для кожного об’єкта $O_i \in O$ ($i = \overline{1, n}$) задано вектор значень ознак $x^{(i)} = (x_{i1}, x_{i2}, \dots, x_{im})$, де x_{ij} – значення ознаки $X_j \in X$ для відповідного об’єкта. Необхідно побудувати правило $s = D(x^{(i)})$, яке здійснює відображення множини $X^{(M)}$ можливих значень ознак $\{X_1, X_2, \dots, X_m\}$ на множину $\{1, 2, \dots, K\}$ номерів кластерів, тобто

$$s: X^M \rightarrow \{1, 2, \dots, K\}. \quad (1)$$

Іншими словами, задача кластеризації полягає у визначенні індикаторної змінної s таким чином:

$$s = \begin{cases} 1, \text{ якщо } x^{(i)} \in \Omega_1, \\ \dots \\ K, \text{ якщо } x^{(i)} \in \Omega_K, \end{cases} \quad (2)$$

де $\Omega_k, (k = \overline{1, K})$ – області простору $X^{(M)}$, які відповідають кластерам.

Для розв’язання задачі застосуємо метод еволюційної кластеризації EvoClast [8], в основі якого лежить генетичний алгоритм, який класично включає такі етапи:

1. Визначення популяції індивідів Θ , які є потенційними розв’язками задачі оптимізації цільової функції.

2. Реалізація попередніх кроків алгоритму, які полягають у визначенні кількості елементів l вибіркової початкової популяції Θ , де $l \ll |\Theta|$; вибір правила нормалізації вхідних даних; вибір способів рекомбінації, мутації та інверсії й відповідних імовірностей.

3. Для кожного елемента $\theta_i \in \Theta$, $i = \overline{1, l}$, обчислюємо значення цільової функції $f_i = F(\theta_i)$.

4. З ймовірностями p_i^l , пропорційними значенням f_i , вибираємо два індивіди та реалізуємо рекомбінація для отримання двох нових індивідів.

5. З ймовірністю $\frac{1}{2}$ вибираємо одного із отриманих індивідів та з ймовірністю P^m виконуємо мутацію.

6. Повторюємо кроки 3–5 $\left\lceil \frac{l}{2} \right\rceil$ рази.

7. Формуємо новий склад популяції Θ , вилучаємо індивіди з гіршими значеннями цільової функції.

Пропонуємо алгоритм побудови цільової функції для розв'язання сформульованої вище задачі.

Припустимо, що всі компоненти векторів ознак об'єктів $x^{(i)} = (x_{i1}, x_{i2}, \dots, x_{im})$ є числовими. Тоді, на початковому етапі здійснюємо нормалізацію даних за формулою

$$x_{ij} := \frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}}. \quad (3)$$

Після здійснення такого перетворення, значення всіх векторів даних будуть міститись в одиничному гіперкубі $[0;1]^m$.

Задаємо величину K – кількість кластерів. Позначимо $T = m \cdot K$ – кількість елементів у вектора-індивіда популяції. Тоді, цільова функція F буде залежати від T змінних: $F = F(y_{11}, y_{12}, \dots, y_{1m}, y_{21}, y_{22}, \dots, y_{2m}, \dots, y_{Km})$, де $(y_{k1}, y_{k2}, \dots, y_{km})$ – центр k -го кластеру на відповідній ітерації ($k = \overline{1, K}$).

Далі послідовно виконуємо такі кроки:

Крок 1. Присвоюємо початкове значення фітнес функції $F := 0$.

Крок 2. Для кожного елементу вибіркової множини O послідовно виконуємо наступні кроки. Нехай $i = 1$.

Крок 3. Обчислюємо відстані від об'єкта O_i до кожного з K кластерів за формулою:

$$d_k = \sqrt{\sum_{j=1}^m (x_{ij} - y_{kj})^2}. \quad (4)$$

Крок 4. Відносимо i -ий об'єкт до q -го кластеру, де $q = \arg \min_{k=\overline{1, K}} d_k$.

Крок 5. Коригуємо значення цільової функції за правилом: $F := F + d_q$. Переходимо до наступного об'єкта: $i := i + 1$.

Крок 6. Якщо були розглянуті всі об'єкти з множини O , тобто $i = n + 1$, то процес обчислення значення функції F завершено, інакше – переходимо до кроку 3.

Вираз для обчислення цільової функції можна записати таким чином:

$$F(y_{11}, \dots, y_{km}) = \sum_{i=1}^n \left(\min_{k=1, K} \left\{ \sqrt{\sum_{j=1}^m (x_{ij} - y_{kj})^2} \right\} \right). \quad (5)$$

Тоді задача кластеризації полягає у знаходження мінімального значення цільової функції (5).

5. Експериментальна верифікація методу еволюційної кластеризації для задачі визначення підгруп ГПР – трудових мігрантів

Для виконання експериментальної верифікації модифікованого методу еволюційної кластеризації розглянуто задачу визначення підгруп ГПР – трудових мігрантів. Вихідними даними в поставленій задачі є результати попередньо проведеного соціологічного дослідження серед трудових мігрантів – жителів Закарпатської області України, в якому проанкетовано 561 представник цільової групи.

Аналіз отриманих анкет проводився за наступними напрямками: соціально-демографічні характеристики респондентів (стать, вік, місцевість постійного проживання, освіта та сімейний стан); напрямок трудової міграції (зовнішня трудова міграція – країни Європи та, окремо, Російська Федерація; внутрішня трудова міграція – інші області України); поведінкові особливості (практики ризикованої щодо інфікування ВІЛ поведінки).

Математично задача виділення підгруп ГПР була представлена як задача кластеризації. Для виконання кластеризації було взято $n = 561$ анкет осіб – трудових мігрантів. За основу виокремлено такі соціально-демографічні ознаки: стать особи, вікова категорія, місце проживання, освіта, сімейний стан та напрямок трудової міграції ($m = 6$). Фрагмент даних наведений в табл. 1.

Кожному значенню якісної ознаки поставлено у відповідність числове значення згідно з табл. 2.

Після переведу якісних ознак в числові аналоги та виконанні нормалізації за формулою (3) проведено кластеризацію за алгоритмом кластеризації k-середніх та описаним в дослідженні еволюційним алгоритмом для різної кількості кластерів ($K = \overline{3,8}$). Для кожного розбиття обчислено значення цільової функції F за формулою (5). Результати обчислень наведено в табл. 3 та на рис. 1.

За даними табл. 3 та рис. 1 значення цільової функції, одержані за методом еволюційної кластеризації є меншими, ніж при застосуванні методу k-середніх, що свідчить про доцільність розв'язання задачі кластеризації групи осіб – трудових мігрантів методом еволюційної кластеризації.

При кількості кластерів $K = 6$ методом еволюційної кластеризації було отримано такі результати, що наведені в табл. 4.

Таблиця 1

Фрагмент соціально-демографічних характеристик осіб – трудових мігрантів

№	Соціально-демографічні ознаки					
	Стать особи	Вікова категорія особи	Місце проживання особи	Освіта особи	Сімейний стан особи	Напрямок міграції особи
...
10	чоловік	3 (35–44)	село	середня спеціальна	одружений / цивільний шлюб	Україна
11	чоловік	2 (25–34)	село	неповна середня	у шлюбі ніколи не перебував	Україна
12	чоловік	2 (25–34)	село	повна середня	одружений / цивільний шлюб	Захід
13	жінка	4 (45 і більше)	село	середня спеціальна	одружений / цивільний шлюб	Захід
14	жінка	2 (25–34)	село	середня спеціальна	у шлюбі ніколи не перебував	Україна
15	чоловік	4 (45 і більше)	місто	повна середня	розлучений / вдовець	Захід
16	чоловік	2 (25–34)	село	неповна середня	одружений / цивільний шлюб	Захід
17	чоловік	2 (25–34)	село	повна середня	одружений / цивільний шлюб	Україна
18	чоловік	2 (25–34)	село	середня спеціальна	одружений / цивільний шлюб	Україна
...

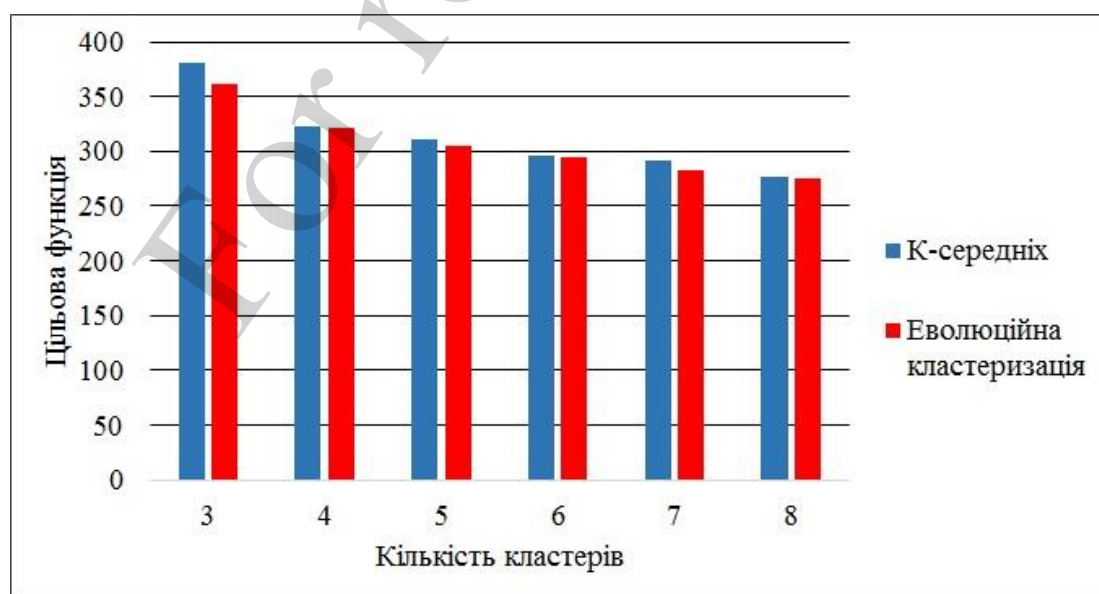


Рис. 1. Діаграма значень цільової функції для різних методів кластеризації

Таблиця 2

Таблиця відповідності значень ознак соціально-демографічного портрету їх числовим аналогам

Соціально-демографічна ознака	Значення ознаки	Числовий аналог
Стать	жінка	0
	чоловік	1
Вікова категорія	1 (18–24)	0
	2 (25–34)	1
	3 (35–44)	2
	4 (45 і більше)	3
Місце проживання	село	0
	місто	1
Освіта	неповна середня	0
	середня	1
	середня спеціальна	
	неповна вища	2
	вища	
Сімейний стан	у шлюбі ніколи не перебував	0
	одружений / цивільний шлюб	1
	розлучений / вдівець	2
Напрямок міграції	Захід	0
	Україна	1
	Росія	2

Таблиця 3

Значення цільової функції

Кількість кластерів (K)	Метод к-середніх (F_1)	Метод еволюційної кластеризації (F_2)
3	380,6709	361,5065
4	322,4295	321,5448
5	311,1273	305,5683
6	295,9276	294,6649
7	290,8294	282,5672
8	276,0774	275,7684

На основі отриманих результатів – центрів кластерів, відповідно до табл. 2 є можливою побудова соціально-демографічного портрету типового представника-кластера за його ознаками. Розподіл по кластерах за соціально-демографічними ознаками представників цільової групи дослідження представлено в табл. 5.

Для подальшого аналізу та інтерпретації результатів кластеризації з поведінкових анкет осіб – трудових мігрантів була сформована система ознак

$\{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$, які визначають практики ризикованої щодо інфікування ВІЛ поведінки цих осіб (табл. 6).

Таблиця 4
Результати кластеризації (центри кластерів)

№ клас- теру	y_{k1}	y_{k2}	y_{k3}	y_{k4}	y_{k5}	y_{k6}
1	0,975	0,481	0,99994	0,773	0,499	0,266
2	0,0013	0,438	0,027	0,546	0,537	0,2219
3	0,99997	0,507	0,0053	0,494	0,473	0,2004
4	0,99984	0,755	0,00703	0,4708	0,5297	0,965
5	0,998	0,254	0,00378	0,478	0,236	0,902
6	0,959	0,452	0,9998	0,556	0,491	0,976

Таблиця 5
Кластери за соціально-демографічними ознаками

Кластер (кількість предста- вників)	Соціально-демографічні ознаки					
	Стать	Вікова катего- рія	Місце прожи- вання	Освіта	Сімейний стан	Напря- мок міг- рації
1 ($n=63$)	чоловік	25-44	місто	середня / вища	одружений / ци- вільний шлюб	Захід / Україна
2 ($n=83$)	жінка	24-34	село	середня	одружений / ци- вільний шлюб	Захід / Україна
3 ($n=168$)	чоловік	25-44	село	середня	одружений / ци- вільний шлюб	Захід / Україна
4 ($n=22$)	чоловік	35 і більше	село	середня	одружений / ци- вільний шлюб	Росія
5 ($n=190$)	чоловік	25-34	село	середня	у шлюбі ніколи не перебував / одружений / ци- вільний шлюб	Росія
6 ($n=35$)	чоловік	25-34	місто	середня	одружений / ци- вільний шлюб	Росія

Наявність у особи ознак $u_1 - u_4$ свідчить про високий поведінковий ризик інфікування ВІЛ. Аналогічно про це ж свідчить відсутність у особи ознак $u_5 - u_7$. Для кожного кластеру обчислені відсотки його представників (трудових мігрантів), для яких ці ознаки мають місце. Результати наведені в табл. 7.

Як представлено в табл. 7, практикують ризиковану поведінку щодо інфікування ВІЛ найменша частка представників другого та третього кластерів. Так, лише 13,2 % представників кластеру 2 відзначили наявність випадкових статевих партнерів (ознака u_1) – це найнижчий показник серед усіх сформова-

них кластерів. Для представників цього кластеру також характерні низькі відсотки представників, що мають ознаки ризикованої поведінки (u_2, u_3, u_4) та відносно висока частка представників, для яких характерні ознаки відсутності ризиків інфікування ВІЛ ($u_5 - u_7$). У той же час відсоток представників кластерів 1 та 4, яким притаманні ознаки ризикованої поведінки ($u_1 - u_4$), є значно вищим.

Таблиця 6

Описові характеристики ознак, які визначають практики ризикованої щодо інфікування ВІЛ поведінки трудових мігрантів

Ознака	Опис ознаки як практики ризикованої щодо інфікування ВІЛ поведінки
u_1	Наявність випадкових статевих партнерів протягом останніх 12 місяців
u_2	Наявність сексуальних відносин з випадковими статевими партнерами в регіоні виїзду на заробітки без використання презервативу
u_3	Наявність у минулому інфекцій, що передаються переважно статевим шляхом
u_4	Наявність сексуальних відносин з випадковим статевим партнером в регіоні постійного проживання без використання презервативу
u_5	Відсутність випадкових статевих партнерів
u_6	Наявність випадкових статевих партнерів протягом останніх 12 місяців, але сексуальні відносини з ними лише з презервативом
u_7	Відсутність сексуальних відносин або усі статеві контакти лише з презервативом

Таблиця 7

Відсотки представників кластерів, яким притаманні ознаки, що визначають практики ризикованої щодо інфікування ВІЛ поведінки

Номер Кластеру	u_1	u_2	u_3	u_4	u_5	u_6	u_7
1	33	17,4	66,6	31	68,2	9	68
2	13,2	10,8	59	12	87,9	9,6	87,9
3	26,1	9,5	50,5	22	77,9	10,1	77,9
4	40,9	0	59	40,9	59	4,5	59
5	24,7	11,5	61	23,1	76,8	11,5	76,8
6	28	22	65,7	25,7	74,2	5,7	74,2

Аналіз результатів дозволив всіх осіб – трудових мігрантів, відповідно до ступеня ризикованості їх поведінки щодо інфікування ВІЛ, розділити на три групи кластерів у порядку зростання ступеня ризику: група 1 – група кластерів відносно низького ризику інфікування ВІЛ, група 2 – група кластерів середнього ризику та

група 3 – група кластерів високого ризику. В результаті такого поділу кожна група кластерів є однорідною не лише за соціально-демографічними портретами її представників, але і за ступенем поширеності практик ризикованої щодо інфікування ВІЛ поведінки. Розподіл кластерів за соціально демографічними ознаками по групах за ступенем ризику інфікування ВІЛ наведений в табл. 8.

Таблиця 8

Групування осіб – трудових мігрантів відповідно до ступеня ризику щодо інфікування ВІЛ

Група ризику	Кластер	Соціально-демографічні ознаки					
		Стать	Вікова категорія	Місце проживання	Освіта	Сімейний стан	Напрямок міграції
1	2, 3	жінка	24–34	село	середня	одружений/цивільний шлюб	Захід
		чоловік	25–44				
2	5, 6	чоловік	25–34	село	середня	у шлюбі ніколи не перебував/одружений/цивільний шлюб	Росія
				місто		одружений/цивільний шлюб	
3	4, 1	чоловік	35 і більше	село	середня	одружений/цивільний шлюб	Росія
			25–44	місто	середня / вища		Україна

Як показано в табл. 8, до групи кластерів високого ризику інфікування ввійшли представники 1 та 4 кластерів, соціально-демографічний портрет яких такий: чоловіки вікової категорії 25–44 роки, що мають закінчену середню або вищу освіту, одружені. При цьому серед жителів села цієї групи осіб найвищий ризик мають трудові мігранти, які виїжджають до Росії, а серед жителів міста – до країн Західної Європи та інших областей України. Представники кластерів 2 та 3, які входять до групи з відносно низьким ризиком інфікування ВІЛ, є жителями сільської місцевості, одружені з закінченою середньою освітою. Це жінки та чоловіки молодого (24–34 роки) та зрілого віку (до 45 років), які виїжджають на заробітки до країн Західної Європи.

Експериментальна верифікація адаптованого методу еволюційної кластеризації для розв'язання задачі формування кластерів осіб – трудових мігрантів довела його ефективність. Перевагою адаптованого методу еволюційної кластеризації для розв'язання поставленої задачі є його більша точність в порівнянні з методом k-середніх.

Використання запропонованого методу дозволило провести уточнену кластеризацію представників групи підвищеного ризику інфікування ВІЛ – трудових мігрантів за основними соціально-демографічними ознаками та сформувати групи кластерів за ознакою ступеню ризику інфікування ВІЛ їх представників. Такий підхід дозволяє при плануванні профілактичних заходів та інтервенцій в програмах протидії епідемії ВІЛ концентрувати увагу та ресурси на найбільш уразливих підгрупах представників ГПР, які є однорідними за своїми соціально-демографічними ознаками. Особливо це актуально в умовах обмеження як фінансових, так людських ресурсах в сфері протидії епідемії ВІЛ, яке наразі спостерігається в Україні.

Доцільним також є застосування адаптованого методу еволюційної кластеризації до аналізу ГПР в інших регіонах України, а також до соціальних груп іншого походження загалом.

6. Висновки

Запропоновано розв'язання задачі визначення ризиків інфікування ВІЛ для окремих групах населення як задачі кластеризації об'єктів. У ході дослідження:

1. Розроблено математичну модель задачі кластеризації як задачі побудови правила, за яким здійснюється відображення з множини можливих значень ознак на множину кластерів та модифіковано метод еволюційної кластеризації об'єктів, використання якого дозволило виконати поділ об'єктів на кластери, застосовуючи еволюційну парадигму. Модифікацію методу здійснено шляхом побудови фітнес-функції у вигляді такої цільової функції, яка забезпечує віднесення об'єкта до того кластеру, евклідова відстань від центру якого до об'єкта є найменшою.

2. Виконано експериментальну верифікацію модифікованого методу еволюційної кластеризації для задачі визначення підгруп за соціально-демографічними ознаками в ГПР – трудових мігрантів, що дозволило здійснити розбиття представників цієї ГПР на кластери відповідно до ступенів ризику інфікування ВІЛ. Сформовано три групи кластерів у порядку зростання ступеня ризику інфікування ВІЛ: група кластерів високого ризику, група кластерів середнього ризику та група відносно низького ризику інфікування ВІЛ. Відповідно, представники даної ГПР, у яких наявні соціально-демографічні ознаки, що відповідають характеристикам кластерів 1 і 4 (група кластерів відносно високого ризику інфікування ВІЛ) потенційно мають більш високий рівень ризику інфікування і, відповідно, потребують ціленаправлених профілактичних заходів з метою попередження інфікування. Застосований підхід дозволяє концентрувати увагу і ресурси на найбільш уразливій категорії осіб даної ГПР, що є основою підвищення ефективності протидії поширенню ВІЛ-інфекції/СНІДу серед представників даної групи населення.

3. Здійснено порівняльний аналіз результатів розв'язання задачі кластеризації заданої ГПР методом k-середніх та методом еволюційної кластеризації. Порівняння здійснено за значеннями функції F , яка є інтегральною сумою відстаней від об'єктів до центрів тих кластерів, яким вони належать. Менше значення цієї функції відповідає кращому розбиттю на кластери. Таким чином,

відповідно до проведених обчислень, було доведено переваги саме еволюційного методу.

Література

1. HIV infection in Ukraine: Information Bulletin number 44 [Text]. – Health of Ukraine, Eng. control center for the socially dangerous diseases. Kyiv, 2015. – 37 p.
2. Harmonised Ukraine report on progress in the implementation of the national response to the AIDS epidemic. Reporting period [Text]. – Ministry of Health of Ukraine, 2012. – 243 p.
3. Grushetsky, A. Monitoring of behavior and HIV prevalence among commercial sex workers as surveillance component of the second generation analytical report on the results of bio-behavioral studies 2011 [Text] / A. Grushetsky // International Alliance HIV/AIDS Alliance in Ukraine, 2012. – 120 p.
4. Bejan, A. B. Standards program correcting deviant behavior of adolescents aged 14–18 years who have experience of substance use [Text] / A. B. Bejan. – Kyiv, 2015. – 180 p.
5. Myronyuk, I. S. Regional high risk of HIV infection in Transcarpathia is workers [Text] / I. S. Myronyuk, G. A. Slabkiy // Ukraine. Health of the Nation. – 2012. – Vol. 3, Issue 23. – P. 201–205.
6. Mulesa, O. Yu. Information technologies of quantitative evaluation of risk groups of human immunodeficiency virus infection [Text] / O. Yu. Mulesa // Eastern-European Journal of Enterprise Technologies. – 2013. – Vol. 5, Issue 4 (65). – P. 10–15. Available at: <http://journals.urau.ua/eejet/article/view/18327/16087>
7. Gorban, A. N. Method of Elastic Maps and its Applications in Data Visualization and Data Modelling [Text] / A. N. Gorban, A. Yu. Zinovyev. // Int. Journal of Computing Anticipatory Systems, CHAOS. – 2002. – Vol. 12. – P. 353–369.
8. Snytyuk, V. Evolutionary clustering of complex objects and processes. Vol. 1 [Text] / V. Snytyuk // XI-th International Conference «Knowledge–Dialogue–Solution», 2005. – P. 232–237.
9. Baturkin, S. A. Statistical data clustering algorithms in adaptive learning systems [Text] / S. A. Baturkin, E. Yu. Baturkina, V. A. Zimenko, Y. V. Sihinov // Vestnyk RHRTU. – 2010. – Vol. 1, Issue 31. – P. 82–85.
10. Fukunaha, K. Introduction to statistical pattern recognition theory [Text] / K. Fukunaha. – Moscow: Nauka. Hlavnaia redaktsiya fizyko–matematycheskoi literatury, 1979. – 368 p.
11. Tsekouras, G. E. On training RBF neural networks using input–output fuzzy clustering and particle swarm optimization [Text] / G. E. Tsekouras, J. Tsimikas // Fuzzy Sets and Systems. – 2013. – Vol. 221. – P. 65–89. doi: 10.1016/j.fss.2012.10.004
12. Horbachenko, V. Y. Networks and Kohonen maps [Electronic resource] / V. Y. Horbachenko. – 2013. – Available at: http://gorbachenko.self-organization.ru/articles/Self-organizing_map.pdf

13. Mundhada, M. R. Record Linkage in Various Types of Clustering Tree By using K-means Clustering Algorithm [Text] / M. R. Mundhada, A. R. Nimje // National Conference "CONVERGENCE", 2015. – P. 28.
14. Smeshko, Yu. V. On a criterion to select the exponential weight classification algorithm of fuzzy C-means [Text] / Yu. V. Smeshko. – Molodezh y nauk, Krasnoyarsk, 2012. – P. 19–25.
15. Mulesa, O. Adaptation of fuzzy c-means method for determination the structure of social groups [Text] / O. Mulesa // Technology Audit and production reserves. – 2015. – Vol. 2, Issue 2 (22). – P. 73–76. doi: 10.15587/2312-8372.2015.41014
16. Santi, É. A model for clustering data from heterogeneous dissimilarities [Text] / É. Santi, D. Aloise, S. J. Blanchard // European Journal of Operational Research. – 2016. – Vol. 253, Issue 3. – P. 659–672. doi: 10.1016/j.ejor.2016.03.033
17. Oliveira, R. M. Hybrid heuristics based on column generation with path-relinking for clustering problems [Text] / R. M. Oliveira, L. A. Lorena, A. A. Chaves, G. R. Mauri // Expert Systems with Applications. – 2014. – Vol. 41, Issue 11. – P. 5277–5284. doi: 10.1016/j.eswa.2014.03.008
18. Bakoben, M. Improving clustering performance by incorporating uncertainty [Text] / M. Bakoben, A. Bellotti, N. Adams // Pattern Recognition Letters. – 2016. – Vol. 77. – P. 28–34. doi: 10.1016/j.patrec.2016.03.004
19. Wu, C.-H. Automated clustering by support vector machines with a local-search strategy and its application to image segmentation [Text] / C.-H. Wu, C.-C. Lai, C.-Y. Chen, Y.-H. Chen // Optik – International Journal for Light and Electron Optics. – 2015. – Vol. 126, Issue 24. – P. 4964–4970. doi: 10.1016/j.ijleo.2015.09.143
20. Carrizosa, E. Clustering categories in support vector machines [Text] / E. Carrizosa, A. Nogales-Gómez, D. R. Morales. – Omega, 2016. doi: 10.1016/j.omega.2016.01.008